ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645

Indian Journal of Science and Technology, *Vol 9(24), 10.17485/ijst/2016/v9i24/96151, June 2016*

A Study on Birth Prediction and BCG Vaccine Demand Prediction using ARIMA Analysis

Keun-Won Kim[1], Guozhong Li[2], Seong-Taek Park[3] and Mi-Hyun Ko[4]*

[1]Department of Management, Sogang University, Korea; dark-kgw@hanmail.net
[2]Department of Management Science and Information System, Kunming University of Science and Technology, China; misgukchung@nate.com
[3]Department of Management Information Systems, Chungbuk National University, Republic of Korea; solpherd@cbnu.ac.kr
[4]Department of Policy Research, Korea Institute of Science and Technology Information, Republic of Korea; mi-hyungo@kisti.re.kr

Abstract

Background/Objectives: This study was conducted to solve the problem by predicting vaccine demand in advance through analysis of progress of birth of newborn babies in our country. **Methods/Statistical Analysis:** The deducted problem was defined and information, data, and use analysis method and planning procedures for creating alternatives for this issue were conducted. Afterwards, R which is an open source analysis tool was used to analysis and for visualization. In this analysis, a time series model (ARIMA model, Box-Jenkins methodology) was used to predict demand and perform the research to predict the number of births in Republic of Korea. **Findings**: The tuberculosis vaccines in Korea are currently being entirely of imported ones. However, the import volume often lacks meeting the demand. In this paper, research was performed to predict the demand of tuberculosis vaccines to secure vaccine stock. As result of analysis, the number of births next year was predicted to be 445,558 (in 2016). Also, analyzed results showed that approximately 388,251 to 502,864 babies will be born in reliability level of 85% and that approximately 357,915 to 533,200 babies will be born in reliability level of 95%. Vaccine should be prepared standard to the minimum value within error range because vaccines have expiration dates. Also, if more births occur than the predicted result, the issue can be coped in prior plans of preparing BCG seal-type vaccines by comparing with monthly predicted number of births. **Application/Improvements:** The results of this study will be applied to the ways to politically solve problems such as supply and demand of BCG vaccine for the expected newborns.

Keywords: ARIMA, Big Data, BCG Vaccine, Demand Prediction, Forecasts

1. Introduction

Great interest has risen in various social classes as the term big data has appeared in daily news media and is becoming a familiar term like the Internet. Big data is a field that is rising as the next-generation growth engine to lead our economic development along with IoT (Internet of Thing) or cloud service[1].

Big data is created and used in several fields in which health and medical fields are receiving most spotlight[2]. Due to the increase of chronic diseases and degenerative diseases according to the aging society, various researches are being attempted to use big data to reduce medical costs, prevent infectious diseases, and improve medical services. In medical organizations, big data technology is being introduced to development of new medicine,

*Author for correspondence

therapy, and diagnosis technologies by using various bio-markers and machine learning algorithms and efficiently saving analyzing the mass data that is accumulated due to the digitization of medical records[3].

Problems on disease management have been pointed out recently due to the MERSC disease crisis. However, tuberculosis is a disease that causes higher death rate than MERSC. According to OECD statistics, mortality due to tuberculosis is shown to be the highest in the world. There are injection-type and seal-type BCG vaccines among tuberculosis vaccines. In Korea, BCG injection-type vaccine is being vaccinated free of charge following the WHO recommendation subject to national essential vaccination. However, demand prediction of BCG injection-type vaccine that is being entirely imported has failed and it is actually unclear when import will be possible due to continuous delay.

However, BCG injection must be vaccinated within 4 weeks of birth that the problem occurred that expensive costs must be spent on BCG seal-type vaccines[4–7]. Prediction on demand is failing every year, but research on this demand issue is not being dealt with.

Therefore, this research was conducted to solve the problem by predicting vaccine demand in advance through analysis of progress of birth of newborn babies in this country.

2. Literature Review

Along with public information opening policies by the government, researches using health and medical field big data are actively being conducted domestically. In[2] researched the risk of cerebrovascular diseases according to change of bio-marker in blood using standard cohort DB which is the big data of the National Health Insurance Corporation. In this research, the influence of change of bio-marker in blood on risk of cerebrovascular diseases were quantitatively evaluated with the goal to investigate the proper period of preventative mediation. As result of the research, the risk of cerebrovascular diseases by change and speed of bio-marker in blood was investigated in which it is seen that prediction of the disease can be enhanced and that the results can be used as basis of setting appropriate inspection period and items of health inspections[8].

In[9] conducted research using the HIRA_NPS (National Patients Sample) provided by the Health Insurance Review and Assessment Service. In this study, the domestic prevalence rate of diabetes in 2009 and aspect of drug prescription were to be grasped using typical sample data. The size of prevalent patients and prevalence rate were each calculated from the sample data and population data, and blood sugar reducing prescription aspect was investigated subject to type 2 diabetes patients. As result, predicted diabetes prevalence rate results using the sample data corresponded to the population analysis results and it was found that the predicted prescription rate by each medicinal effect of blood sugar reducing prescription also corresponded the population analysis results[9].

These researches using big data in the health and medical field are being conducted, but there is almost no research related to tuberculosis. In[10] used the Delphi method to perform basic investigation on national public vaccine R&D policy establishment. However, the opinions of several experts went through several feed-backs using self-administered questionnaire or mail survey using standardized and nonstandardized tools to converge and meet agreement of opinion using the Delphi method in this study rather than analysis using data[10].

Previous researches using health and medical field big data have been conducted, but there is almost no research on tuberculosis. In this study, big data is used to perform research on the prediction of vaccine demand.

3. Research Method

In this study, the deducted problem was defined and information, data, and use analysis method planning procedures for creating alternatives for this issue were conducted. Afterwards, R which is an open source analysis tool was used to attempt analysis and visualization to progress the study by analyzing these results[11].

3.1 Problem Definition

According to OECD statistics, mortality due to tuberculosis is shown to be the highest in the world. To prevent tuberculosis, it has been selected as a national vaccination target in our country in which BCG injection-type vaccine is vaccinated free of charge. However, lack of vaccine demand occurs every year due to the failure of predicting

demand that expensive costs are spent on BCG seal-type vaccine when free vaccination is not available. To solve this issue, data was analyzed to predict demand.

3.2 Data Collection

To conduct this research, the 'Annual Report of Current Condition of 2013 Tuberculosis Patients' provided by the Ministry of Health and Welfare, 'Mortality by Main Cause of Death per 100,000 Population by OECD Nations' and 'Monthly, Annual Population Trend (birth, death, marriage, divorce statistics)' statistics provided by the National Statistics Portal were used to perform the research.

4. Data Analysis

First, mortality due to tuberculosis was investigated according to OECD nations. This is a bar graph of mortality due to tuberculosis per 100,000 population by OECD nations standard to 2012.

As shown in Figure 1, it can be known that Republic of Korea has the highest mortality. In 2012, the mortality due to tuberculosis in ROK was 4.4% which is highest in the world. Korea was followed by Mexico (2.6%) with 1.7 times higher mortality and has 44 times higher mortality than USA (0.1%).

Despite this fact, tuberculosis vaccines are still being imported. BCG vaccine in our country is currently being entirely imported from Denmark.

Figure 2 shows the number of new tuberculosis patients of each year in our country. As shown in the figure, new patients occur every year. 2014 shows a decrease compared to 2013, but it can be seen in a large picture that tuberculosis patients are gradually increasing.

Annual increase is shown, but lack of demand occurs every year due to the failure of predicting vaccine demand. According to this data, birth in our country is to be predicted through ARIMA analysis and vaccine demand is to be predicted based on this result.

4.1 Analysis Model

In this analysis, a time series model (ARIMA model) was used to predict demand and perform the research to predict the number of births in our country.

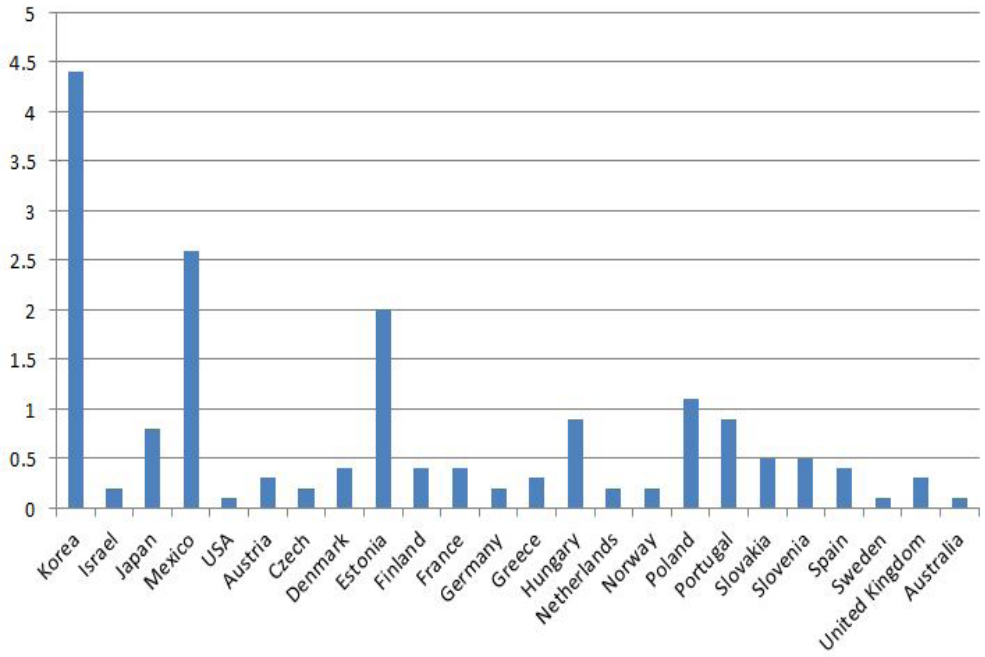

Figure 1. Mortality (tuberculosis) according to OECD nations in 2012.

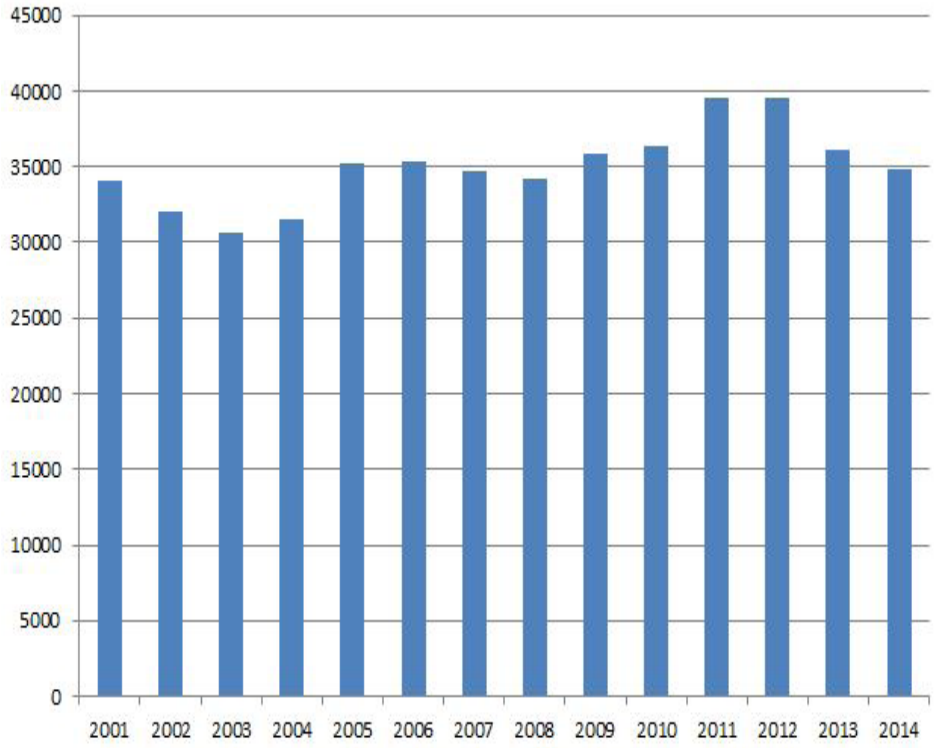

Figure 2. Annual number of new tuberculosis patients in ROK (2001-2014).

4.1.1 Concept of Time Series Analysis

Time series analysis is a method used to analyze a property value by grasp characteristics only with current and past values without considering other variables that have causal relationship with that property value. Thus, analysis takes place only based on the past form or measured value of that variable when predicting a future variable in the time series analysis.

The nonstructural approach method that is the basis of this time series model received theoretical support from the neoclassical school related to the currency principle. However, there were several limitations and restrictions in the method of analysis, but substantial parts could be supplemented by the development of computer software development on problems that have large scale structural equations. After the 1980s, this was commonly used in economic analysis and prediction duty to lead great development.

4.1.2 Meaning of ARIMA Model

4.1.2.1 Probabilistic Time Series Model

In the time series analysis, future measured values of a variable on time series data, thus continuous time are predicted in which the ARIMA (autoregressive integrated moving average model) model designed by Box and Jenkins is commonly used. It must be assumed that time series data is created under probabilistic assumption to analyze and apply this model that the model created under this probabilistic assumption must be a probabilistic model.

The AR process model and MA process model are representative probabilistic models and the ARMA (autoregressive moving average model) model is used when the probabilistic procedure has both autoregressive process and autoregressive moving average process.

4.1.2.2 ARIMA Model

The models shown above are based on the assumption of stable time series, but most economic time series are unstable and these can be changed into stable processes through one or two differences. If the time series wt gained by difference of the unstable Yt by d times becomes a stable time series, we call Yt the homogeneous non-stationary process of order "d".

The time series stabilized through difference can be expressed in an AR model, MA model or ARMA model and the stable time series wt (=) gained by d times difference of Yt can be expressed into an ARMA (p, q) model which is a normal type that the original time series Yt is called the ARIMA (p, d, q) model, thus integrated ARMA process of order (p, d, q).

4.1.3 Box-Jenkins Methodology

Recent time series analyses have rapidly developed that plans to predict more systematic and effective future values can be prepared in which this is called the Box-Jenkins methodology.

Thus, the method proposed by Box-Jenkins is used to find optimal models when particular time series data exists.

It is a method using the ARIMA (autoregressive integrated moving average) model called the Box-Jenkins model and is composed of 4 steps. However, time series analysis by the ARIMA model is subject to analyzing contents that change according to one value according to time that it has dynamic analysis characteristics in which there is a problem that satisfactory analysis is possible only when at least more than 50 measured values exist[12].

5. Results

Figure 3 shows the result of number of births in our country visualized by the ARIMA analysis. Analysis was conducted using data on the number of births between January 2000 and July 2015. The reason the data was based on monthly standard instead of annual standard is because of the expiration date of BCG vaccine. There was an article that vaccination of BCG injection-type vaccine will be difficult this August.

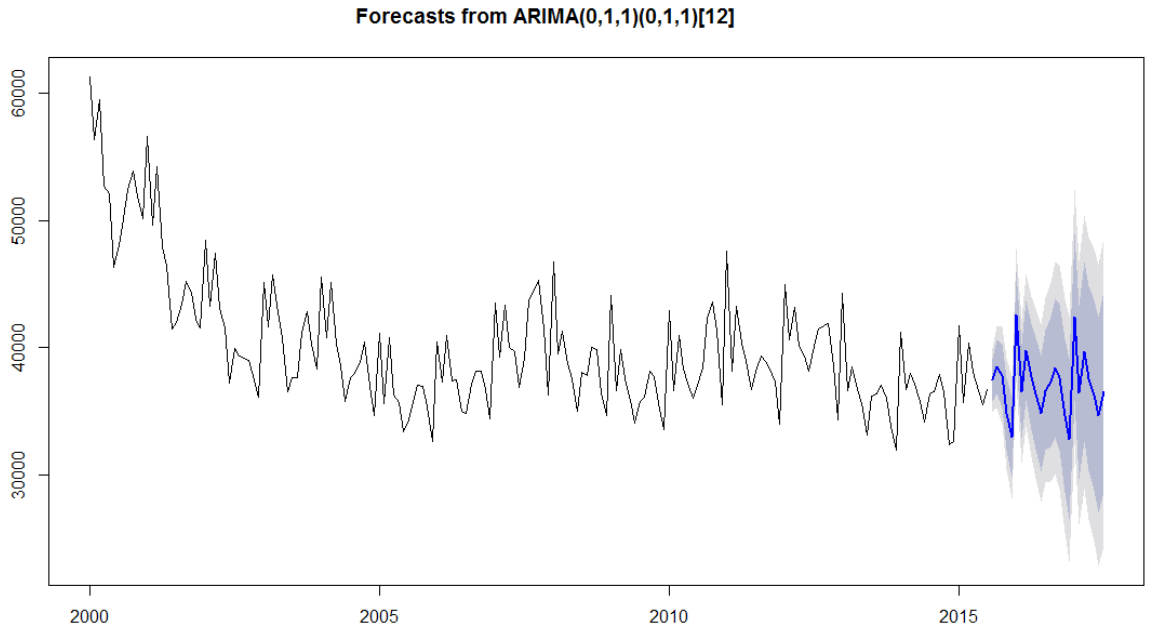

Figure 3. Prediction graph of number of births.

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
Aug 2015	37458.49	35827.20	39089.79	34963.65	39953.34
Sep 2015	38517.55	36405.72	40629.37	35287.79	41747.31
Oct 2015	37820.49	35318.79	40322.18	33994.47	41646.50
Nov 2015	34817.21	31978.69	37655.73	30476.07	39158.35
Dec 2015	32957.96	29818.55	36097.36	28156.65	37759.27
Jan 2016	42549.25	39135.38	45963.13	37328.18	47770.33
Feb 2016	36590.91	32923.05	40258.78	30981.40	42200.43
Mar 2016	39799.14	35893.77	43704.51	33826.39	45771.89
Apr 2016	37735.61	33606.37	41864.85	31420.49	44050.74
May 2016	36419.49	32077.92	40761.07	29779.62	43059.37
Jun 2016	34843.53	30299.53	39387.54	27894.08	41792.99
Jul 2016	36672.94	31935.15	41410.74	29427.12	43918.77
Aug 2016	37333.54	32218.58	42448.50	29510.89	45156.19
Sep 2016	38392.60	32962.35	43822.84	30087.76	46697.43
Oct 2016	37695.53	31967.33	43423.73	28935.01	46456.06
Nov 2016	34692.26	28680.85	40703.67	25498.60	43885.91
Dec 2016	32833.00	26551.14	39114.87	23225.72	42440.28

Figure 4. ARIMA analysis results.

It was told in the article that the expiration date of the BCG injection-type vaccine that our country secures was until September 2nd and that quick vaccination is required beforehand. Also, newborn babies who could not receive free vaccination due to lack of vaccine stock had to be vaccinated with expensive BCG seal-type vaccine. That is why the monthly number of births was predicted to perform the analysis to predict the demand of required vaccine.

Figure 4 shows the results of predicted number of births from August 2015 to December 2016 using ARIMA analysis. As shown in the figure, monthly predicted numbers of births are indicated. Also, predicted error values are shown. Predicted result values of reliability level of 80% and 95% are shown.

Predicted results showed that the total number of births was predicted to be 445,558 in 2016. Also, analyzed results showed that approximately 388,251 to 502,864 babies will be born in reliability level of 85% and that approximately 357,915 to 533,200 babies will be born in reliability level of 95%.

It is considered that BCG vaccine issues that occur every year can be prepared in advance if BCG vaccine is imported by predicting the number of annual births by using this analysis model.

6. Conclusion and Limitations

In this study, the number or births in 2016 was predicted using the number or births between January 2000 and July 2015. Through these analyzed results, it was predicted that 445,558 babies will be born in 2016. Based on the results gained from the analysis with 95% reliability level, it is predicted that at least 357,915 BCG injection-type vaccines must be imported in 2016.

Vaccines have expiration dates that vaccines must be prepared standard to the minimum value of predicted error range. Also, if more births occur than predicted results or if vaccine is lacked compared to the predicted monthly number of births, problems can be solved by plans such as preparing BCG seal-type vaccine. Vaccine demand issues that occur every year and difficulty to predict vaccine due to diseases such as MERSC have occurred last year. However, vaccine demand should be prepared beforehand through prediction if the number of births can be predicted to some level.

As vaccination is especially essential to babies within 1 month of birth, vaccine preparation and quick measures should be prepared through prediction. This study has limitation at the fact that only the number of births was used to predict the number of newborn babies in 2016. It is considered that better predicted values can be gained

if the number of newborn babies is predicted by considering the ratio of adult men and women, marriage rate, and population by age. Also, it is considered that issues such as vaccine crisis can be solved if vaccine demand is grasped in prior with these results.

7. References

1. Oh SG. Health insurance medical advantage of big data. Research Institute for Healthcare Policy Korean Medical Association. 2015; 12(3):18–23.
2. Kim BS, Kim DY, Kim KW, Park ST. The improvement plan for fire response time using big data. Indian Journal of Science and Technology. 2013 Sep; 8(23):1–5.
3. Lee JH, Jae MY, Cho MG, Son HS. Leverage big data trends in the healthcare sector. The Journal of The Korean Institute of Communication Sciences. 2014 Dec; 32(1):63–75.
4. Public Health TB vaccine, from next month, 2 days vaccination crisis. Available from: http://news20.busan.com/controller/newsController.jsp?newsId=20150828000124
5. The another came vaccine chaos. Available from: http://www.hankyung.com/news/app/newsview.php?aid=2015082813571
6. Vaccine chaos. Available from: http://www.doctorsnews.co.kr/news/articleView.html?idxno=106441
7. BCG vaccine supply followed by stop in this year. Available from: http://news20.busan.com/controller/newsController.jsp?newsId=20151016000100
8. Kim HC. Cerebrovascular disease risk based on changes in serum biomarkers. Big Data Pilot Study Published Symposium; 2013.
9. Park BJ. Korea, diabetes prevalence estimates and DPP-4 inhibitors using assessment aspects. Health Insurance Review and Assessment Service National Patient sample data (HIRA-NPS: National Pastients Sample) using a Symposium; 2011.
10. Lee SM, Yeo SG, Kang SJ, Han SY, Lee SW. A delphi study on national public vaccine research and development policy in Korea. Health Policy and Management. 2015; 25(2):140–8.
11. Kim DW, Kang TG, Li G, Park ST. Analysis of user's behaviors and growth factors of shopping mall using big-data. Indian Journal of Science and Technology. 2013 Oct; 8(25):1–7.
12. Kim DH. Analysis of the price forecasts in housing market with ARIMA model. Journal of Korea Real Estate Society. 2014 Dec; 32(2):277–94.